

(translation of the front page of the priority document of  
Japanese Patent Application No. 2000-378019)



PATENT OFFICE  
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the  
following application as filed with this Office.

Date of Application: December 12, 2000

Application Number : Patent Application 2000-378019

Applicant(s) : Canon Kabushiki Kaisha

November 2, 2001

Commissioner,  
Patent Office

Kouzo OIKAWA

Certification Number 2001-3096834

日本国特許庁  
JAPAN PATENT OFFICE

09/972,996  
Kenichi Fujii  
October 10, 2001

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application:

2000年12月12日

出願番号

Application Number:

特願2000-378019

出願人

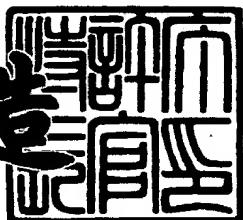
Applicant(s):

キヤノン株式会社

2001年11月 2日

特許庁長官  
Commissioner,  
Japan Patent Office

及川耕造



【書類名】 特許願  
【整理番号】 4255026  
【提出日】 平成12年12月12日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 3/00  
【発明の名称】 音声認識システム及び音声認識装置ならびにそれらの方法  
【請求項の数】 19  
【発明者】  
【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内  
【氏名】 清水 智之  
【発明者】  
【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内  
【氏名】 池田 裕治  
【特許出願人】  
【識別番号】 000001007  
【氏名又は名称】 キヤノン株式会社  
【代理人】  
【識別番号】 100076428  
【弁理士】  
【氏名又は名称】 大塚 康徳  
【電話番号】 03-5276-3241  
【選任した代理人】  
【識別番号】 100101306  
【弁理士】  
【氏名又は名称】 丸山 幸雄  
【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100115071

【弁理士】

【氏名又は名称】 大塚 康弘

【電話番号】 03-5276-3241

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0001010

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識システム及び音声認識装置ならびにそれらの方法

【特許請求の範囲】

【請求項1】 第1の装置と第2の装置がネットワーク内に通信可能に接続された音声認識を行うシステムであって、

前記第1の装置が、

音声を入力する1つ以上の音声入力手段と、該音声入力手段により入力された音声を認識する音声認識手段とを備え、

前記第2の装置が、

前記音声入力手段により入力された音声を認識する音声認識手段を備え、

前記第1の装置は、前記音声入力手段により入力された音声を、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを明示的に切り替える切替手段を備えることを特徴とする音声認識システム。

【請求項2】 前記切替手段は、グラフィカル・ユーザ・インターフェースを提供するG U I入出力手段からの指示入力を受け付ける切替指示受付手段を更に備え、該切替指示受付手段が受け付けた指示入力に従って、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを切り替えることを特徴とする請求項1に記載の音声認識システム。

【請求項3】 前記切替手段は、前記音声入力手段から入力された特定の音声を指示入力として受け付ける音声指示受付手段を更に備え、該音声指示受付手段が受け付けた指示入力に従って、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを切り替えることを特徴とする請求項1に記載の音声認識システム。

【請求項4】 前記切替手段は、1つ以上の前記音声入力手段のうち、特定の音声入力手段からの音声入力の有無を指示入力として受け付ける音声入力特定手段を更に備え、該音声入力特定手段が受け付けた指示入力に従って、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを切り替えることを特徴とする請求項1に記載の音声認識システム。

【請求項5】

前記第1の装置はクライアントとして機能し、前記第2の装置はサーバとして機能することを特徴とする請求項1に記載の音声認識システム。

【請求項6】 音声を入力する1つ以上の音声入力手段と、

前記音声入力手段により入力された音声を認識する音声認識手段と、

前記音声入力手段により入力された音声を、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを明示的に切り替える切替手段と、

前記切替手段にて、前記第2の装置の音声認識手段で音声認識することを選択した場合に、前記音声入力手段にて入力された音声データを該第2の装置へ送信する送信手段と、

前記第2の装置で音声認識された認識結果を受信する受信手段とを備えることを特徴とする音声認識装置。

【請求項7】 前記切替手段は、グラフィカル・ユーザ・インターフェースを提供するG U I入出力手段からの指示入を受け付ける切替指示受付手段を更に備え、該切替指示受付手段が受け付けた指示入力に従って、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを切り替えることを特徴とする請求項6に記載の音声認識装置。

【請求項8】 前記切替手段は、前記入力手段から入力された特定の音声を指示入として受け付ける音声指示受付手段を更に備え、該音声指示受付手段が受け付けた指示入力に従って、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを切り替えることを特徴とする請求項6に記載の音声認識装置。

【請求項9】 前記切替手段は、1つ以上の前記音声入力手段のうち、特定の音声入力手段からの音声入力の有無を指示入として受け付ける音声入力特定手段を更に備え、該音声入力特定手段が受け付けた指示入力に従って、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを切り替えることを特徴とする請求項6に記載の音声認識装置。

【請求項10】 第1の装置と第2の装置がネットワーク内に通信可能に接続された音声認識を行うシステムにおける音声認識方法であって、

前記第1の装置において、1つ以上の音声入力手段を用いて音声入力し、入力された音声を認識する音声認識工程と、

前記第2の装置において、前記音声入力手段により入力された音声を認識する音声認識工程と、

前記第1の装置において、前記音声入力手段により入力された音声を、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを明示的に切り替える切替工程とを備えることを特徴とする音声認識方法。

【請求項11】 前記切替工程は、グラフィカル・ユーザ・インターフェースを提供するGUI入出力手段からの指示入力を受け付ける切替指示受付工程を更に備え、該切替指示受付工程が受け付けた指示入力に従って、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを切り替えることを特徴とする請求項10に記載の音声認識方法。

【請求項12】 前記切替工程は、前記音声入力手段から入力された特定の音声を指示入力として受け付ける音声指示受付工程を更に備え、該音声指示受付工程が受け付けた指示入力に従って、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを切り替えることを特徴とする請求項10に記載の音声認識方法。

【請求項13】 前記切替工程は、1つ以上の前記音声入力手段を用いて音声入力し、特定の音声入力手段からの音声入力の有無を指示入力として受け付ける音声入力特定工程を更に備え、該音声入力特定工程が受け付けた指示入力に従って、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを切り替えることを特徴とする請求項10に記載の音声認識方法。

【請求項14】 前記ネットワークにおいて、

前記第1の装置はクライアントとして機能し、前記第2の装置はサーバとして機能することを特徴とする請求項10に記載の音声認識方法。

【請求項15】 1つ以上の音声入力手段を用いて音声入力し、入力された

音声を認識する音声認識工程と、

前記音声入力手段により入力された音声を、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを明示的に切り替える切替工程と、

前記切替工程にて、前記第2の装置で実行される音声認識工程で音声認識することを選択した場合に、前記音声入力手段にて入力された音声データを該第2の装置へ送信する送信工程と、

前記第2の装置で音声認識された認識結果を受信する受信工程とを備えることを特徴とする音声認識方法。

【請求項16】 前記切替工程は、グラフィカル・ユーザ・インターフェースを提供するG U I入出力手段からの指示入を受け付ける切替指示受付工程を更に備え、該切替指示受付工程が受け付けた指示入力に従って、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを切り替えることを特徴とする請求項15に記載の音声認識方法。

【請求項17】 前記切替工程は、前記音声入力手段から入力された特定の音声を指示入力として受け付ける音声指示受付工程を更に備え、該音声指示受付工程が受け付けた指示入力に従って、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを切り替えることを特徴とする請求項15に記載の音声認識方法。

【請求項18】 前記切替工程は、1つ以上の前記音声入力手段を用いて音声を入力し、特定の音声入力手段からの音声入力の有無を指示入力として受け付ける音声入力特定工程を更に備え、該音声入力特定工程が受け付けた指示入力に従って、前記第1の装置で実行される音声認識工程と前記第2の装置で実行される音声認識工程のいずれで音声認識するかを切り替えることを特徴とする請求項15に記載の音声認識方法。

【請求項19】 請求項10～18のいずれか1つに記載の音声認識方法をコンピュータによって実現させるための制御プログラムを格納する記憶媒体。

【発明の詳細な説明】

## 【0001】

## 【発明の属する技術分野】

本発明は、クライアント・サーバ型音声認識システムにおいて、クライアントとサーバ両者で音声認識が可能な場合に、ユーザから入力された音声データをどちらで認識するかを切り替えることが可能な音声認識システムに関するものである。

## 【0002】

## 【従来の技術】

例えば、非力な携帯端末で音声認識機能を実現させる場合などに、負荷の大きな音声認識を十分な計算機資源と性能を持つ外部の音声認識サーバに分担させる、いわゆるクライアント・サーバ型の音声認識システムが利用されている（特開平7-222248）。

## 【0003】

一方で、クライアント・サーバ型音声認識システムを使用する場合、大語彙などの高度な認識を行う場合や、ネットワークに接続して使用する前提のあるものはサーバ側で音声認識を行うが、クライアント側の操作のような少数語彙の音声認識をクライアント上で行うことで、ネットワークのトラフィックを軽減できるようにした音声入力クライアント装置が考えられてきている。

## 【0004】

## 【発明が解決しようとする課題】

従来のシステムにおいては、入力された音声データからはクライアント装置とサーバ装置のどちらの音声認識手段が使用されているかが不明瞭で、判断がつかないという問題がある。また、クライアント装置とサーバ装置で同一の語彙を認識辞書に持っていた場合にも同様の問題が起る。

## 【0005】

本発明は、上記の課題に鑑みてなされたものであり、入力された音声データをクライアント装置とサーバ装置のどちらで認識するかを明示的に切り替えることで、クライアント装置に音声入力した場合に発生する、音声認識手段切り替えの不明瞭性を解決することを目的とする。

【0006】

【課題を解決するための手段】

上記の課題を解決するために本発明の音声認識システムは以下のような構成を備える。すなわち、

第1の装置と第2の装置がネットワーク内に通信可能に接続された音声認識を行なうシステムであって、

前記第1の装置が、

音声を入力する1つ以上の音声入力手段と、該音声入力手段により入力された音声を認識する音声認識手段とを備え、

前記第2の装置が、

前記音声入力手段により入力された音声を認識する音声認識手段を備え、

前記第1の装置は、前記音声入力手段により入力された音声を、前記第1の装置の音声認識手段と前記第2の装置の音声認識手段のいずれで音声認識するかを明示的に切り替える切替手段を備える。

【0007】

【発明の実施の形態】

以下、図面を参照して本発明の一実施形態を詳細に説明する。

【0008】

図1は、本発明の一実施形態に係る音声認識システムの基本構成を示すブロック図である。

【0009】

同図において、101はユーザの入力音声から音声認識部104および105で認識可能な音声データを生成する音声入力部である。102は音声入力部101で生成した音声データを切り替え指示受信部103の指示により音声認識部を音声入力クライアント装置の音声認識部104ないしサーバ装置の音声認識部105のいずれかに切り替える音声認識先切り替え部である。切り替え指示受信部103は、音声入力クライアント装置の音声認識部ないしはサーバ装置の音声認識部105のいずれかを使用するように切り替える指示を受信し、音声認識先切り替え部102に切り替え指示を伝える。104、105は音声入力部101で

生成した音声データを認識する音声認識部である。この時104は音声入力クライアント装置上に存在し、105はサーバ装置上に存在する。

## 【0010】

図2は、本発明の一実施形態に係る音声入力クライアント装置のシステム構成を示すブロック図である。

## 【0011】

同図の音声入力クライアント装置200において、201は音声入力部101で処理する音声を獲得するための例えばマイクなどの音声入力装置である。202はユーザの切り替え指示が入力される例えばボタンやキーなどの音声入力クライアント上の物理的な切り替え指示入力装置である。203はサーバ装置との間で音声入力クライアント装置がデータの送受信を行う為の通信装置である。204は図1における各処理を行うプログラムを格納するROMである。205はROM204に格納されたプログラム実行時のワーク用のRAMである。206はROM204およびRAM205に格納したプログラムを実行するCPUである。207は本装置の各部を接続してデータをやり取りするバスである。208は音声入力クライアント装置とサーバ装置を接続してデータを送受信するためのネットワークである。209はクライアント・サーバ型音声認識システムにおける音声認識可能なサーバ装置であり本実施形態においては1つである。サーバ装置209の詳細については省略する。

## 【0012】

次に図3に示すフローチャートを用いて本実施形態における音声認識システムの概要を説明する。

## 【0013】

まず、ステップS301では、音声入力クライアント装置200の音声入力装置201に対して、ユーザから入力された音声を音声入力部101が音声認識部104、105で認識できる音声データ形式にして取得する。ステップS302では、切り替え指示受信部103が切り替え指示入力装置202に対するユーザからの、例えばボタンを押す等の切り替え指示が入力されているかを監視し、切り替え指示を受信している場合はステップS303に移る。切り替え指示を受信

していない場合はステップS305に移る。ステップS303では、音声認識先切り替え部102がステップS301で取得した音声データを音声入力クライアント装置200で音声認識するように音声データを送信する。ステップS304ではS303で送信されてきたデータを音声入力クライアント装置の音声認識部104で認識する。そして認識結果の処理へと続くが以降の詳細は省略する。

#### 【0014】

また、ステップS305では切り替え指示が出されていない為、音声認識先切り替え部102は送信先を切り替えず、通常のグライアント・サーバ型音声認識システム同様、音声データをサーバ装置209へ送信し、サーバ装置209の音声認識部105で認識を行う。認識結果は、音声入力クライアント装置200に返信され、音声入力クライアント装置200にて、例えば表示、印刷或いはコマンドとして出力される（ステップS306）。なお、ステップS306での処理は一般的なクライアント・サーバ型の音声認識システムに準ずるため、ここでは詳細は省略する。

#### 【0015】

##### 【他の実施形態】

上記実施形態においては、ユーザが使用する入力装置202として物理的なボタンを想定したが、この限りではない。例えば、図4のようなシステム構成をとることも可能である。すなわち、音声入力クライアント装置400に図2における入力装置202に替えてG U I入出力装置402を設けることによってグラフィカル・ユーザ・インターフェース（G U I）をユーザに提供する。そして、図1の切り替え指示受信部103においてG U Iによる切り替え指示入力を処理し、これによって音声認識先の切り替えを行っても良い。

#### 【0016】

##### 【他の実施形態】

上記実施形態においては、ユーザが使用する入力装置202として物理的なボタンを想定したが、この限りではない。例えば、図5に示す基本構成をとることも可能である。すなわち、音声入力部101で生成した音声データを切り替え指示受信部503-1で受信する。このとき、受信した音声データが切り替え指示

を表す特定の音声入力であることを切り替え指示音声認識部503-2が認識した場合に、音声認識切り替え部102に切り替え信号を送信し、音声認識切り替え部102では、切り替え信号にしたがって、それ以降の音声認識を音声入力クライアント装置かサーバ装置のいずれかで行うように切り替えてよい。

## 【0017】

## 【他の実施形態】

上記実施形態においては、ユーザが使用する入力装置202として音声入力クライアント上のボタンを想定したが、この限りではない。例えば、外部のリモコンなどからの無線信号を受信する装置としても良い。

## 【0018】

## 【他の実施形態】

上記実施形態においては、ステップS301において音声認識部104、105で認識できる音声データ形式に変換しているが、例えば、サーバ装置とクライアント装置で音声認識に使用するデータ形式が異なる場合は、音声認識先切り替え部102において音声認識先が決定したあと、その音声認識部で認識可能な音声データ形式に変換しても良い。

## 【0019】

## 【他の実施形態】

上記実施形態においては、切り替え指示をボタンでのみ行うことを想定したが、同時に複数の切り替え指示手段を持っても良い。例えば、ボタン、G.U.I.のどちらを用いても切り替えることが可能とする等である。

## 【0020】

## 【他の実施形態】

上記実施形態においては、いずれもサーバ装置は1つであることを想定しているが、複数のサーバ装置がネットワーク内に存在しても良い。この場合、音声認識先切り替え部は音声入力クライアント装置のものを含め、切り替え指示受信部の指示に従って、使用する音声認識手段を切り替えることも可能である。

## 【0021】

## 【他の実施形態】

上記実施形態においては、切り替え指示を受信することを想定したが、これを設けなくても良い。例えば図6に示すシステム構成をとることも可能である。すなわち、音声入力クライアント装置600に図2における入力装置202に替えて、音声入力装置602を用意する。このとき、図7に示すような音声入力特定部703を用いて音声入力が為された音声入力装置を特定する。これにより、音声入力部A701に対する入力の場合は、音声入力クライアント装置600の音声認識部704で音声認識を行い、音声入力装置B702に対する入力の場合は、サーバ装置上の音声認識部705で音声認識を行う等のようにして切り替えることが可能となる。また、サーバ装置が複数存在する場合は、音声入力装置は必要な個数分さらに複数存在しても良い。

#### 【0022】

##### 【他の実施形態】

上記実施形態においては、音声入力クライアント装置、サーバ装置各々の各部を同一の計算機上で構成する場合について説明したが、これに限定されるものではなく、ネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

#### 【0023】

##### 【他の実施形態】

上記実施形態においては、プログラムをROMに保持する場合について説明したが、これに限定されるものではなく、任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

#### 【0024】

##### 【他の実施形態】

なお、本発明は、複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に適用してもよい。前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはCPUやMPU）が記録媒体に格納されたプログラムコードを読み出し実行することによっても、達成されるることは言うまでもない。この場合、記録媒体から読み出されたプログラムコード自

体が前述した実施形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。

#### 【0025】

##### 【他の実施形態】

プログラムコードを供給するための記録媒体としては、たとえば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、DVD-ROM、DVD-RAM、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

#### 【0026】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOSなどが実際の処理の一部または全部を行ない、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

#### 【0027】

更に、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

#### 【0028】

##### 【発明の効果】

以上説明したように、本発明によれば、入力された音声データをクライアント装置とサーバ装置のどちらで認識するかを明示的に切り替えることで、クライアント装置に音声入力した場合に発生する、音声認識手段切り替えの不明瞭性を解決する効果が得られる。

##### 【図面の簡単な説明】

##### 【図1】

本発明に係る音声入力クライアント装置の実施形態の基本構成を示すブロック図である。

【図2】

本発明に係る音声入力クライアント装置の実施形態のシステム構成を示すブロック図である。

【図3】

本発明に係る音声入力クライアント装置の実施形態の動作を説明するフローチャートである。

【図4】

本発明に係る音声入力クライアント装置に表示装置を持たせた場合の実施形態のシステム構成を示すブロック図である。

【図5】

本発明に係る音声入力クライアント装置に切り替え指示音声認識部を持たせた場合の実施形態の基本構成を示すブロック図である。

【図6】

本発明に係る音声入力クライアント装置に音声入力装置を複数持たせた場合の実施形態の基本構成を示すブロック図である。

【図7】

本発明に係る音声入力クライアント装置に音声入力装置を複数持たせた場合の実施形態の基本構成を示すブロック図である。

【符号の説明】

- 101 音声入力部
- 102 音声認識先切り替え部
- 103 切り替え指示受信部
- 104 音声認識部（クライアント上）
- 105 音声認識部（サーバ上）
- 201 音声入力装置
- 202 入力装置
- 203 通信装置

204 ROM

205 RAM

206 CPU

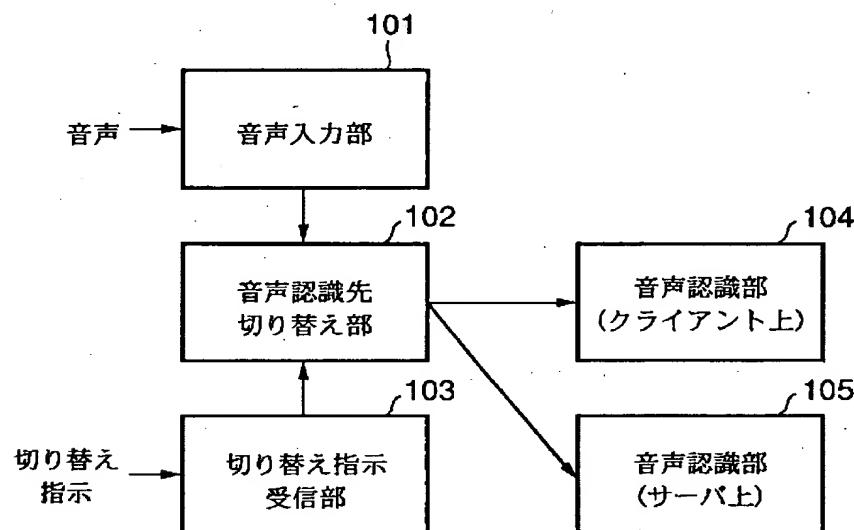
207 バス

208 ネットワーク

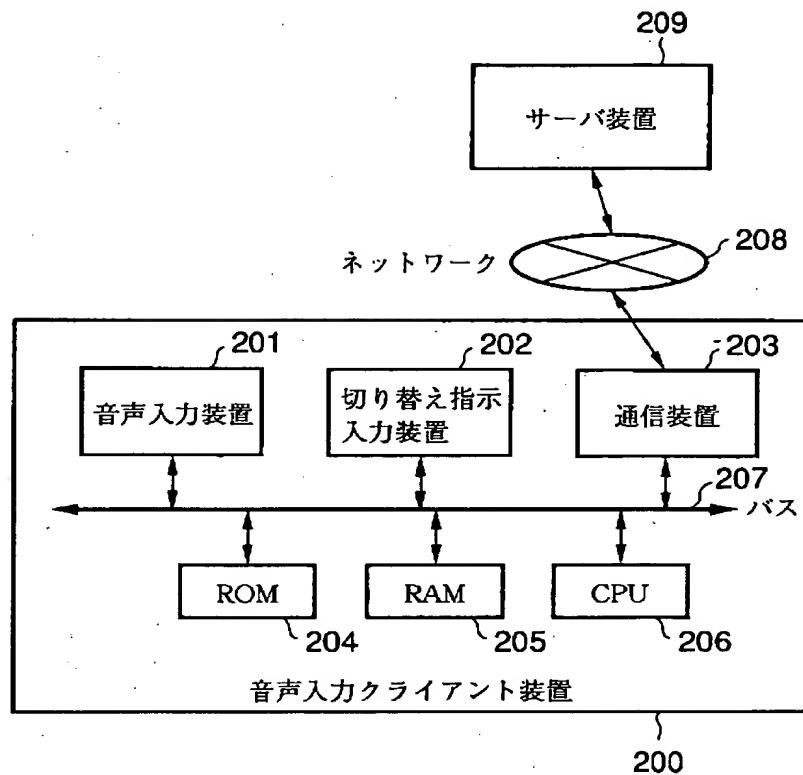
209 サーバ装置

【書類名】 図面

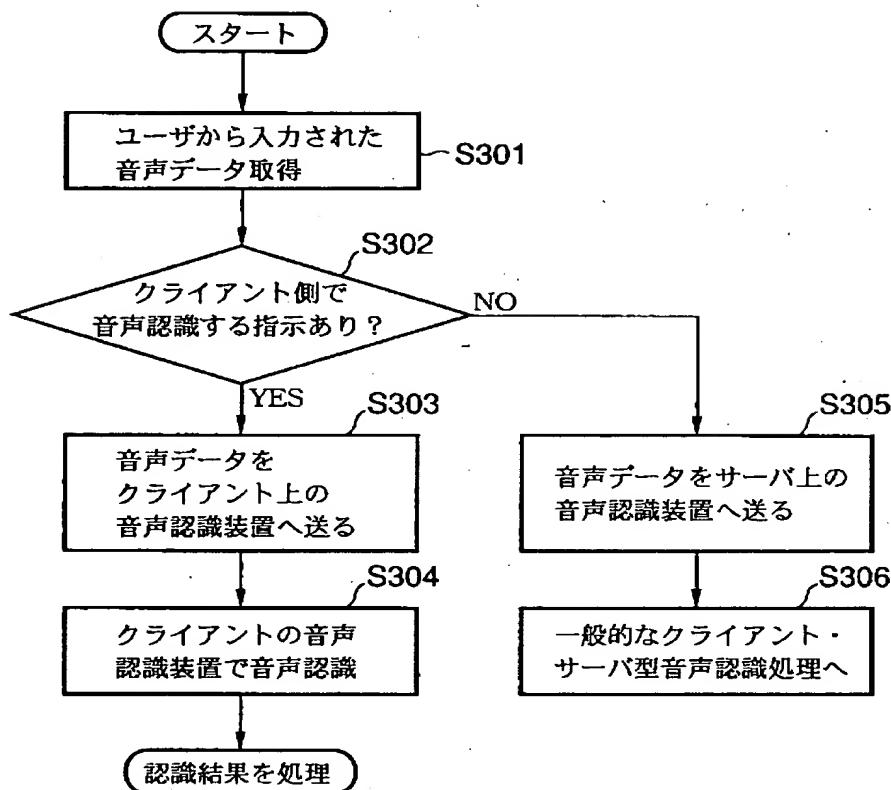
【図1】



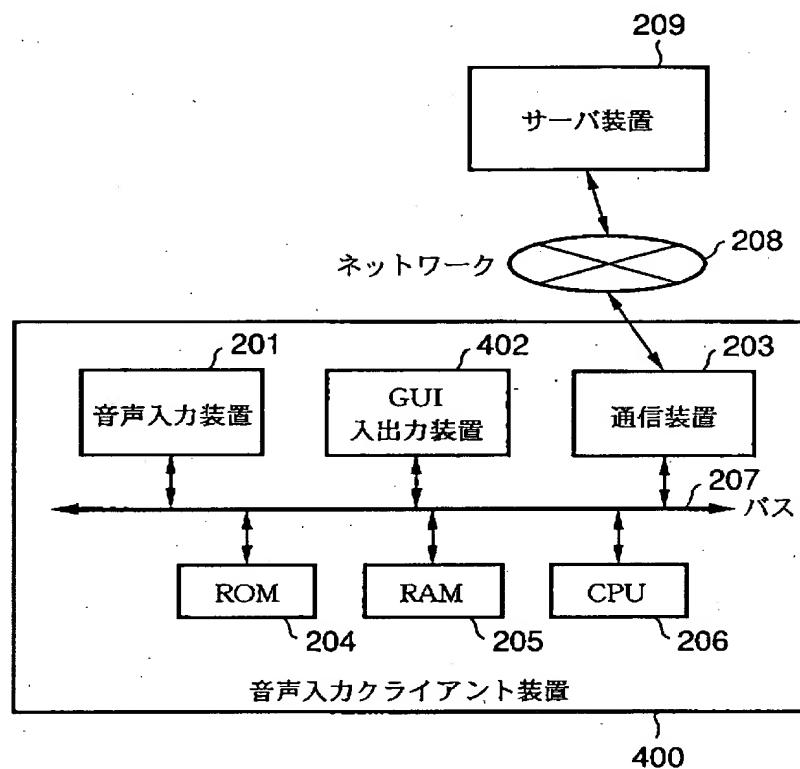
【図2】



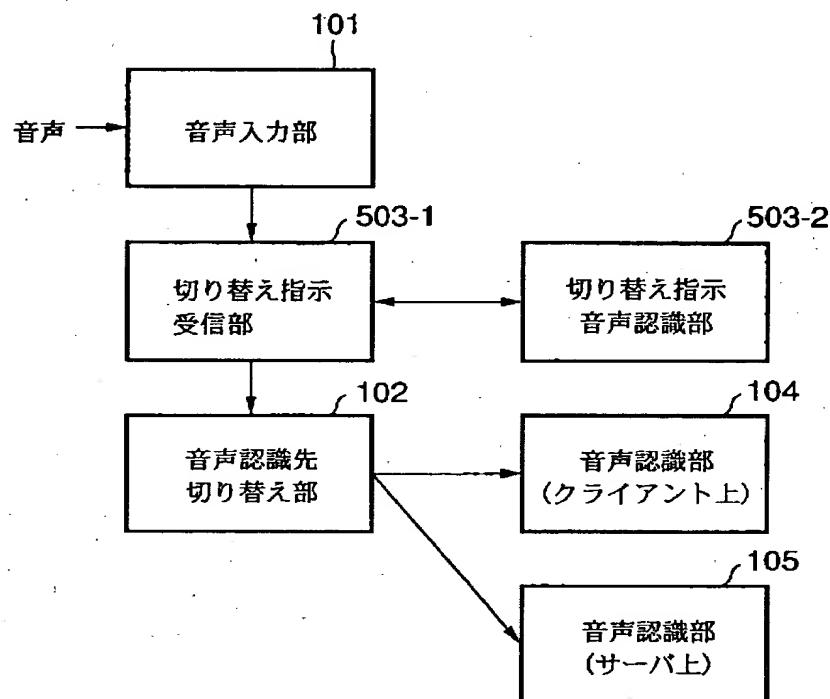
【図3】



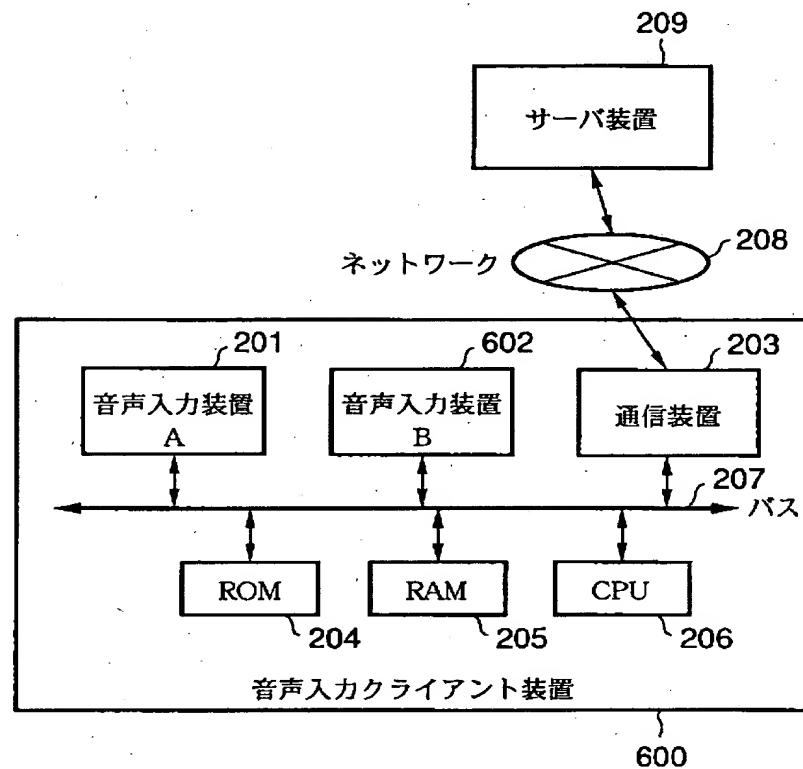
【図4】



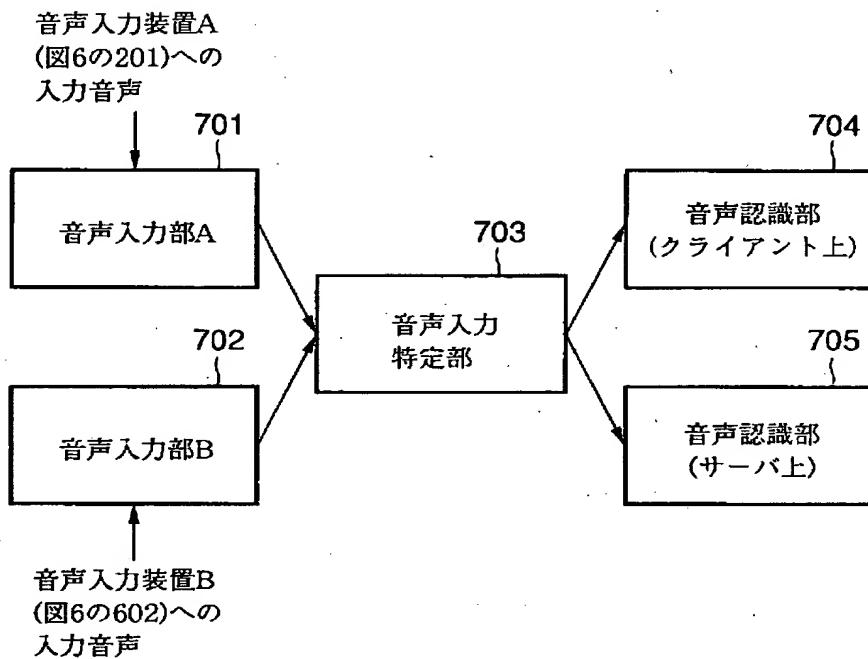
【図5】



【図6】



【図7】



【書類名】 要約書

【要約】

【課題】 クライアント装置に音声入力した場合に発生する、音声認識手段切り替えの不明瞭性を解決する。

【解決手段】 第1の装置と第2の装置がネットワーク内に通信可能に接続された音声認識を行うシステムであって、前記第1の装置が、音声を入力する1つ以上の音声入力部101と、音声入力部101により入力された音声を認識する音声認識部104とを備え、前記第2の装置が、音声入力部101により入力された音声を認識する音声認識部105を備え、前記第1の装置は、音声入力部101により入力された音声を、前記第1の装置の音声認識部104と前記第2の装置の音声認識部105のいずれで音声認識するかを明示的に切り替える音声認識先切り替え部102を備える。

【選択図】 図1

出願人履歴情報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都大田区下丸子3丁目30番2号

氏 名 キヤノン株式会社